

Question 1

What is the optimal value of alpha for ridge and lasso regression? What will be the changes in the model if you choose double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?

Answers: The optimal value of alpha for Ridge is 2 and for Lasso it is 0.001.

Along with alphas the R² in the model near to 0.83. if I doubled the alpha values for Ridge and Lasso, the prediction accuracy come around 0.82 observed negligible change in the co-efficient values. Adapted these changes in model found under question section in jupyter note book

Refer snapshots

Ridge

Ridge Co-Efficient		Ridge Doubled Alpha Co-Efficient	
Total_sqr_footage	0.169122	Total_sqr_footage	0.149028
GarageArea	0.101585	GarageArea	0.091803
TotRmsAbvGrd	0.067348	TotRmsAbvGrd	0.068283
OverallCond	0.047652	OverallCond	0.043303
LotArea	0.043941	LotArea	0.038824
CentralAir_Y	0.032034	Total_porch_sf	0.033870
LotFrontage	0.031772	CentralAir_Y	0.031832
Total_porch_sf	0.031639	LotFrontage	0.027526
Neighborhood_StoneBr	0.029093	Neighborhood_StoneBr	0.026581
Alley_Pave	0.024270	OpenPorchSF	0.022713
OpenPorchSF	0.023148	MSSubClass_70	0.022189
MSSubClass_70	0.022995	Alley_Pave	0.021672
RoofMatl_WdShngl	0.022586	Neighborhood_Veenker	0.020098
Neighborhood_Veenker	0.022410	BsmtQual_Ex	0.019949
SaleType_Con	0.022293	KitchenQual_Ex	0.019787
HouseStyle_2.5Unf	0.021873	HouseStyle_2.5Unf	0.018952
PavedDrive_P	0.020160	MasVnrType_Stone	0.018388
KitchenQual_Ex	0.019378	PavedDrive_P	0.017973
LandContour_HLS	0.018595	RoofMatl_WdShngl	0.017856
SaleType_Oth	0.018123	PavedDrive_Y	0.016840

Lasso

Lasso Co-Efficient		Lasso Doubled Alpha Co-Efficient	
Total_sqr_footage	0.202244	Total_sqr_footage	0.204642
GarageArea	0.110863	GarageArea	0.103822
TotRmsAbvGrd	0.063161	TotRmsAbvGrd	0.064902
OverallCond	0.046686	OverallCond	0.042168
LotArea	0.044597	CentralAir_Y	0.033113
CentralAir_Y	0.033294	Total_porch_sf	0.030659
Total_porch_sf	0.028923	LotArea	0.025909
Neighborhood_StoneBr	0.023370	BsmtQual_Ex	0.018128
Alley_Pave	0.020848	Neighborhood_StoneBr	0.017152
OpenPorchSF	0.020776	Alley_Pave	0.016628
MSSubClass_70	0.018898	OpenPorchSF	0.016490
LandContour_HLS	0.017279	KitchenQual_Ex	0.016359
KitchenQual_Ex	0.016795	LandContour_HLS	0.014793
BsmtQual_Ex	0.016710	MSSubClass_70	0.014495
Condition1_Norm	0.015551	MasVnrType_Stone	0.013292
Neighborhood_Veenker	0.014707	Condition1_Norm	0.012674
MasVnrType_Stone	0.014389	BsmtCond_TA	0.011677
PavedDrive_P	0.013578	SaleCondition_Partial	0.011236
LotFrontage	0.013377	LotConfig_CulDSac	0.008776
PavedDrive_Y	0.012363	PavedDrive_Y	0.008685

Question 2

You have determined the optimal value of lambda for ridge and lasso regression during the assignment. Now, which one will you choose to apply and why?

Answers: The optimal value of alpha for Ridge is 2 and for Lasso it is 0.001

The MSE of the model for Ridge is 0.0018396090787924262

The MSE of the model for Ridge is 0.0018634152629407766

As seen MSE for both the methods are almost same, Lasso helps in feature reduction (as the coefficient value of some of the features become zero), Lasso has a better edge over Ridge and should be used as the final model.

Question 3

After building the model, you realised that the five most important predictor variables in the lasso model are not available in the incoming data. You will now have to create another model excluding the five most important predictor variables. Which are the five most important predictor variables now?

Answers: The five most important predictor variables in the current lasso model is

1. Total_sqr_footage
2. GarageArea
3. TotRmsAbvGrd
4. OverallCond
5. LotArea

build a Lasso model in the Jupiter notebook after removing these attributes from the dataset. The R2 of the new model without the top 5 predictors drops to .73 The Mean Squared Error increases to 0.0028575670906482538

The new Top 5 predictors

Lasso Co-Efficient	
LotFrontage	0.146535
Total_porch_sf	0.072445
HouseStyle_2.5Unf	0.062900
HouseStyle_2.5Fin	0.050487
Neighborhood_Veenker	0.042532

Question 4

How can you make sure that a model is robust and generalisable? What are the implications of the same for the accuracy of the model and why?

Answers: Robustness of a model implies, the testing error of the model is consistent with the training error, the model performs well with enough stability even after adding some noise to the dataset. Thus, the robustness (or generalizability) of a model is a measure of its successful application to data sets other than the one used for training and testing.

By the implementing regularization techniques, can control the trade-off between model complexity and bias which is directly connected the robustness of the model.

Regularization helps in penalizing the coefficients for making the model too complex; thereby allowing only the optimal amount of complexity to the model. It helps in controlling the robustness of the model by making the model optimal simpler. Therefore, in order to make the model more robust and generalizable, one need to make sure that there is a delicate balance between keeping the model simple and not making it too naïve to be of any use. Also, making a model simple leads to Bias Variance Trade-off:

A complex model will need to change for every little change in the dataset and hence is very unstable and extremely sensitive to any changes in the training data.

A simpler model that abstracts out some pattern followed by the data points given is unlikely to change wildly even if more points are added or removed. Bias helps you quantify, how accurate is the model likely to be on test data. A complex model can do an accurate job prediction provided there has to be enough training data.

Models that are too naïve, for e.g., one that gives same results for all test inputs and makes no discrimination whatsoever has a very large bias as its expected error across all test inputs are very high. Variance is the degree of changes in the model itself with respect to changes in the training data. Thus, accuracy of the model can

be maintained by keeping the balance between Bias and Variance as it minimizes the total error as shown in the below graph.

